

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2003-271172
 (43)Date of publication of application : 25.09.2003

(51)Int.Cl.

G10L 13/08
 A63H 5/00
 A63H 11/00
 G10L 13/00
 G10L 13/06
 G10L 21/04

(21)Application number : 2002-073384
 (22)Date of filing : 15.03.2002

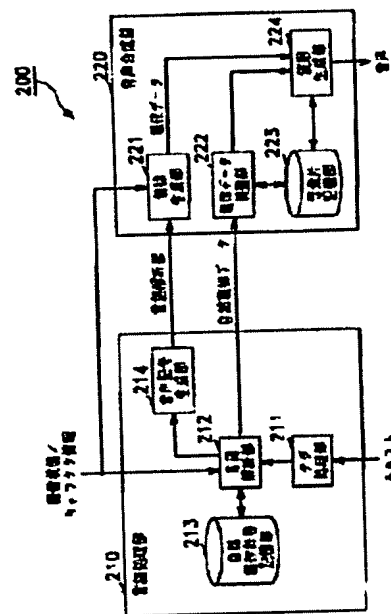
(71)Applicant : SONY CORP
 (72)Inventor : KOBAYASHI KENICHIRO
 YAMAZAKI NOBUHIDE
 AKAHA MAKOTO

(54) METHOD AND APPARATUS FOR VOICE SYNTHESIS, PROGRAM, RECORDING MEDIUM
 AND ROBOT APPARATUS

(57)Abstract:

PROBLEM TO BE SOLVED: To synthesize a natural voice close to a human voice having various expressions such as emotional expressions.

SOLUTION: In a voice synthesis device 200, a natural metrical data dictionary of a natural metrical data dictionary storing part 213 preregisters various words or sentences and their natural metrical data. As for a part registered in the natural metrical data dictionary, corresponding natural metrical data is selected as a result of language processing of an inputted text, and a metrical data adjusting part 222 adjusts a parameter on the basis of phoneme piece data of a phoneme piece storing part 223. Meanwhile, as to a part unregistered in the natural metrical data dictionary, a voice symbol generating part 214 generates a voice symbol string, and a metrical data generating part 221 subsequently generates metrical data. Then, a waveform generating part 224 connects needed phoneme piece data on the basis of the metrical data and the natural metrical data to generate voice waveform data.



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2003-271172
(P2003-271172A)

(43) 公開日 平成15年9月25日 (2003.9.25)

(51) Int.Cl. ⁷	識別記号	F I	テーマコード* (参考)
G 1 0 L 13/08		A 6 3 H 5/00	C 2 C 1 5 0
A 6 3 H 5/00		11/00	Z 5 D 0 4 5
11/00		G 1 0 L 3/00	H
G 1 0 L 13/00			J
13/06			Q

審査請求 未請求 請求項の数41 O L (全 19 頁) 最終頁に続く

(21) 出願番号 特願2002-73384(P2002-73384)

(22) 出願日 平成14年3月15日 (2002.3.15)

(71) 出願人 000002185

ソニー株式会社

東京都品川区北品川6丁目7番35号

(72) 発明者 小林 賢一郎

東京都品川区北品川6丁目7番35号 ソニ

ー株式会社内

(72) 発明者 山崎 信英

東京都品川区北品川6丁目7番35号 ソニ

ー株式会社内

(74) 代理人 100067736

弁理士 小池 晃 (外2名)

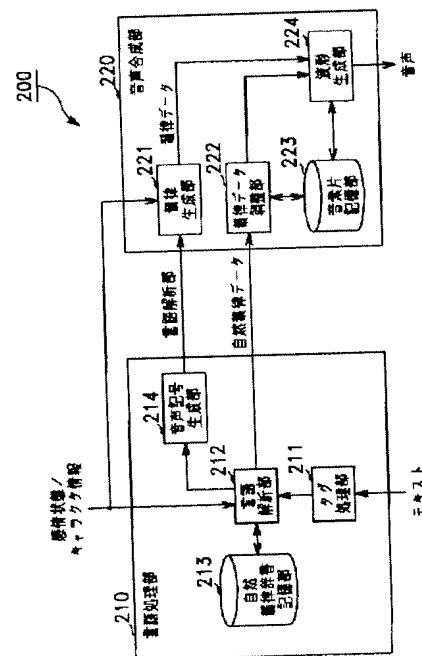
最終頁に続く

(54) 【発明の名称】 音声合成方法、音声合成装置、プログラム及び記録媒体、並びにロボット装置

(57) 【要約】

【課題】 感情表現など多様な表情を持った、人間の声に近い自然な音声合成する。

【解決手段】 音声合成装置200において、自然韻律辞書記憶部213の自然韻律辞書には、予め種々の単語又は文とその自然韻律データとが登録される。入力されたテキストの言語処理の結果、自然韻律辞書に登録されている部分については、対応する自然韻律データを選択し、韻律データ調整部222において音素片記憶部223の音素片データに基づいて、そのパラメータを調整する。一方、自然韻律辞書に登録されていない部分については、音声記号生成部214で音声記号列を生成してから、韻律生成部221で韻律データを生成する。そして、波形生成部224において、韻律データ及び自然韻律データに基づいて必要な音素片データを接続し、音声波形データを生成する。



【特許請求の範囲】

【請求項 1】 入力されたテキストを解析し、当該テキスト中に所定の単語又は文が存在する場合に、予め人間の発声から抽出された上記所定の単語又は文に対応する自然韻律データを記憶手段から選択する自然韻律データ選択工程と、

上記所定の単語又は文以外の部分について、上記テキストの音声記号列を作成する音声記号列作成工程と、
上記音声記号列から韻律データを作成する韻律データ作成工程と、

上記自然韻律データ又は上記韻律データに基づいて音声を合成する音声合成工程とを有することを特徴とする音声合成方法。

【請求項 2】 上記記憶手段には、同一表記に対して複数種類の自然韻律データが記憶されており、
上記自然韻律データ生成工程では、上記複数種類の自然韻律データから、選択制御情報によって指定された 1 種類が選択されることを特徴とする請求項 1 記載の音声合成方法。

【請求項 3】 上記複数種類の自然韻律データは、異なる種類の感情又はキャラクタを表すものであることを特徴とする請求項 2 記載の音声合成方法。

【請求項 4】 上記選択制御情報は、外部から与えられた情報であることを特徴とする請求項 2 記載の音声合成方法。

【請求項 5】 上記選択制御情報は、上記テキストに埋め込まれたタグの情報であることを特徴とする請求項 2 記載の音声合成方法。

【請求項 6】 上記選択制御情報によって指定された自然韻律データが上記記憶手段に記憶されていない場合、
上記自然韻律データ生成工程では、当該指定された自然韻律データに類似する自然韻律データが選択されることを特徴とする請求項 2 記載の音声合成方法。

【請求項 7】 上記音声合成工程では、少なくとも上記自然韻律データのパラメータが調整されて、上記音声合成されることを特徴とする請求項 1 記載の音声合成方法。

【請求項 8】 上記テキストは、複数の音韻を有して成り、上記パラメータは、上記音韻のピッチ周期、継続時間長及び音量の少なくとも 1 つであることを特徴とする請求項 7 記載の音声合成方法。

【請求項 9】 上記音声合成工程では、当該合成に用いられる音素片データの平均ピッチ周期、平均話速及び平均音量の少なくとも 1 つに基づいて、上記パラメータが調整されることを特徴とする請求項 8 記載の音声合成方法。

【請求項 10】 上記記憶手段には、同一表記に対して複数種類の自然韻律データが記憶されており、
上記自然韻律データ生成工程では、上記複数種類の自然韻律データから、選択制御情報によって指定された 1 種類

類が選択され、

上記韻律データ作成工程では、上記選択制御情報に基づいて、上記韻律データの上記パラメータが調整され、
上記音声合成工程では、当該調整後の上記韻律データの上記パラメータに基づいて、上記自然韻律データの上記パラメータが調整されることを特徴とする請求項 8 記載の音声合成方法。

【請求項 11】 所定の単語又は文と、予め人間の発声から抽出された上記所定の単語又は文に対応する自然韻律データとが記憶されている記憶手段と、
入力されたテキストを解析し、当該テキスト中に上記所定の単語又は文が存在する場合に、当該所定の単語又は文に対応する自然韻律データを上記記憶手段から選択する自然韻律データ選択手段と、

上記所定の単語又は文以外の部分について、上記テキストの音声記号列を作成する音声記号列作成手段と、
上記音声記号列から韻律データを作成する韻律データ作成手段と、

上記自然韻律データ又は上記韻律データに基づいて音声を合成する音声合成手段とを備えることを特徴とする音声合成装置。

【請求項 12】 上記記憶手段には、同一表記に対して複数種類の自然韻律データが記憶されており、
上記自然韻律データ生成手段は、上記複数種類の自然韻律データから、選択制御情報によって指定された 1 種類を選択することを特徴とする請求項 11 記載の音声合成装置。

【請求項 13】 上記複数種類の自然韻律データは、異なる種類の感情又はキャラクタを表すものであることを特徴とする請求項 12 記載の音声合成装置。

【請求項 14】 上記音声合成手段は、少なくとも上記自然韻律データのパラメータを調整して、上記音声を合成することを特徴とする請求項 11 記載の音声合成装置。

【請求項 15】 上記テキストは、複数の音韻を有して成り、上記パラメータは、上記音韻のピッチ周期、継続時間長及び音量の少なくとも 1 つであることを特徴とする請求項 14 記載の音声合成装置。

【請求項 16】 上記音声合成手段は、当該合成に用いられる音素片データの平均ピッチ周期、平均話速及び平均音量の少なくとも 1 つに基づいて、上記パラメータを調整することを特徴とする請求項 15 記載の音声合成装置。

【請求項 17】 上記記憶手段には、同一表記に対して複数種類の自然韻律データが記憶されており、
上記自然韻律データ生成手段は、上記複数種類の自然韻律データから、選択制御情報によって指定された 1 種類を選択し、

上記韻律データ作成手段は、上記選択制御情報に基づいて、上記韻律データの上記パラメータを調整し、

上記音声合成手段は、当該調整後の上記韻律データの上記パラメータに基づいて、上記自然韻律データの上記パラメータを調整することを特徴とする請求項15記載の音声合成装置。

【請求項18】 コンピュータに所定の処理を実行させるためのプログラムであって、

入力されたテキストを解析し、当該テキスト中に所定の単語又は文が存在する場合に、予め人間の発声から抽出された上記所定の単語又は文に対応する自然韻律データを記憶手段から選択する自然韻律データ選択工程と、
上記所定の単語又は文以外の部分について、上記テキストの音声記号列を作成する音声記号列作成工程と、
上記音声記号列から韻律データを作成する韻律データ作成工程と、

上記自然韻律データ又は上記韻律データに基づいて音声を作成する音声合成工程とを有することを特徴とするプログラム。

【請求項19】 上記記憶手段には、同一表記に対して複数種類の自然韻律データが記憶されており、
上記自然韻律データ生成工程では、上記複数種類の自然韻律データから、選択制御情報によって指定された1種類が選択されることを特徴とする請求項18記載のプログラム。

【請求項20】 上記複数種類の自然韻律データは、異なる種類の感情又はキャラクタを表すものであることを特徴とする請求項19記載のプログラム。

【請求項21】 上記音声合成工程では、少なくとも上記自然韻律データのパラメータが調整されて、上記音声を作成されることを特徴とする請求項18記載のプログラム。

【請求項22】 上記テキストは、複数の音韻を有して成り、上記パラメータは、上記音韻のピッチ周期、継続時間長及び音量の少なくとも1つであることを特徴とする請求項21記載のプログラム。

【請求項23】 上記音声合成工程では、当該合成に用いられる音声素片データの平均ピッチ周期、平均話速及び平均音量の少なくとも1つに基づいて、上記パラメータが調整されることを特徴とする請求項22記載のプログラム。

【請求項24】 上記記憶手段には、同一表記に対して複数種類の自然韻律データが記憶されており、
上記自然韻律データ生成工程では、上記複数種類の自然韻律データから、選択制御情報によって指定された1種類が選択され、
上記韻律データ作成工程では、上記選択制御情報に基づいて、上記韻律データの上記パラメータが調整され、
上記音声合成工程では、当該調整後の上記韻律データの上記パラメータに基づいて、上記自然韻律データの上記パラメータが調整されることを特徴とする請求項22記載のプログラム。

【請求項25】 コンピュータに所定の処理を実行させるためのプログラムが記録されたコンピュータ読み取り可能な記録媒体であって、

入力されたテキストを解析し、当該テキスト中に所定の単語又は文が存在する場合に、予め人間の発声から抽出された上記所定の単語又は文に対応する自然韻律データを記憶手段から選択する自然韻律データ選択工程と、
上記所定の単語又は文以外の部分について、上記テキストの音声記号列を作成する音声記号列作成工程と、
上記音声記号列から韻律データを作成する韻律データ作成工程と、

上記自然韻律データ又は上記韻律データに基づいて音声を作成する音声合成工程とを有することを特徴とするプログラムが記録された記録媒体。

【請求項26】 上記記憶手段には、同一表記に対して複数種類の自然韻律データが記憶されており、
上記自然韻律データ生成工程では、上記複数種類の自然韻律データから、選択制御情報によって指定された1種類が選択されることを特徴とする請求項25記載の記録媒体。

【請求項27】 上記複数種類の自然韻律データは、異なる種類の感情又はキャラクタを表すものであることを特徴とする請求項26記載の記録媒体。

【請求項28】 上記音声合成工程では、少なくとも上記自然韻律データのパラメータが調整されて、上記音声を作成されることを特徴とする請求項25記載の記録媒体。

【請求項29】 上記テキストは、複数の音韻を有して成り、上記パラメータは、上記音韻のピッチ周期、継続時間長及び音量の少なくとも1つであることを特徴とする請求項28記載の記録媒体。

【請求項30】 上記音声合成工程では、当該合成に用いられる音声素片データの平均ピッチ周期、平均話速及び平均音量の少なくとも1つに基づいて、上記パラメータが調整されることを特徴とする請求項29記載の記録媒体。

【請求項31】 上記記憶手段には、同一表記に対して複数種類の自然韻律データが記憶されており、
上記自然韻律データ生成工程では、上記複数種類の自然韻律データから、選択制御情報によって指定された1種類が選択され、

上記韻律データ作成工程では、上記選択制御情報に基づいて、上記韻律データの上記パラメータが調整され、
上記音声合成工程では、当該調整後の上記韻律データの上記パラメータに基づいて、上記自然韻律データの上記パラメータが調整されることを特徴とする請求項28記載の記録媒体。

【請求項32】 供給された入力情報に基づいて動作を行う自律型のロボット装置であって、

所定の単語又は文と、予め人間の発声から抽出された上

記所定の単語又は文に対応する自然韻律データとが記憶されている記憶手段と、

発話するテキストを解析し、当該テキスト中に上記所定の単語又は文が存在する場合に、当該所定の単語又は文に対応する自然韻律データを上記記憶手段から選択する自然韻律データ選択手段と、

上記所定の単語又は文以外の部分について、上記テキストの音声記号列を作成する音声記号列作成手段と、上記音声記号列から韻律データを作成する韻律データ作成手段と、

上記自然韻律データ又は上記韻律データに基づいて音声を合成する音声合成手段とを備えることを特徴とするロボット装置。

【請求項33】 上記記憶手段には、同一表記に対して複数種類の自然韻律データが記憶されており、上記自然韻律データ生成手段は、上記複数種類の自然韻律データから、選択制御情報によって指定された1種類を選択することを特徴とする請求項32記載のロボット装置。

【請求項34】 上記複数種類の自然韻律データは、異なる種類の感情又はキャラクタを表すものであることを特徴とする請求項33記載のロボット装置。

【請求項35】 上記動作に起因する感情モデルと、上記感情モデルの感情状態を判別する感情状態判別手段とをさらに備え、上記選択制御情報は、上記感情状態判別手段によって判別された感情状態情報であることを特徴とする請求項33記載のロボット装置。

【請求項36】 上記選択制御情報は、上記テキストに埋め込まれたタグの情報であることを特徴とする請求項33記載のロボット装置。

【請求項37】 上記選択制御情報によって指定された自然韻律データが上記記憶手段に記憶されていない場合、

上記自然韻律データ生成手段は、当該指定された自然韻律データに類似する自然韻律データを選択することを特徴とする請求項33記載のロボット装置。

【請求項38】 上記音声合成手段は、少なくとも上記自然韻律データのパラメータを調整して、上記音声を合成することを特徴とする請求項32記載のロボット装置。

【請求項39】 上記テキストは、複数の音韻を有して成り、上記パラメータは、上記音韻のピッチ周期、継続時間長及び音量の少なくとも1つであることを特徴とする請求項38記載のロボット装置。

【請求項40】 上記音声合成手段は、当該合成に用いられる音声素片データの平均ピッチ周期、平均話速及び平均音量の少なくとも1つに基づいて、上記パラメータを調整することを特徴とする請求項39記載のロボット装置。

【請求項41】 上記記憶手段には、同一表記に対して複数種類の自然韻律データが記憶されており、上記自然韻律データ生成手段は、上記複数種類の自然韻律データから、選択制御情報によって指定された1種類を選択し、

上記韻律データ作成手段は、上記選択制御情報に基づいて、上記韻律データの上記パラメータを調整し、上記音声合成手段は、当該調整後の上記韻律データの上記パラメータに基づいて、上記自然韻律データの上記パラメータを調整することを特徴とする請求項39記載のロボット装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、人間の声に近い自然な音声を合成するための音声合成方法、音声合成装置、プログラム及び記録媒体、並びに音声を出力するロボット装置に関する。

【0002】

【従来の技術】電気的又は磁気的な作用を用いて人間（生物）の動作に似た運動を行う機械装置を「ロボット」という。我が国においてロボットが普及し始めたのは、1960年代末からであるが、その多くは、工場における生産作業の自動化・無人化等を目的としたマニピュレータや搬送ロボット等の産業用ロボット（Industrial Robot）であった。

【0003】最近では、人間のパートナーとして生活を支援する、すなわち住環境その他の日常生活上の様々な場面における人的活動を支援する実用ロボットの開発が進められている。このような実用ロボットは、産業用ロボットとは異なり、人間の生活環境の様々な局面において、個々に個性の相違した人間、又は様々な環境への適応方法を自ら学習する能力を備えている。例えば、犬、猫のように4足歩行の動物の身体メカニズムやその動作を模した「ペット型」ロボット、或いは、2足直立歩行を行う人間等の身体メカニズムや動作をモデルにしてデザインされた「人間型」又は「人間形」ロボット（Humanoid Robot）等のロボット装置は、既に実用化されつつある。

【0004】これらのロボット装置は、産業用ロボットと比較して、エンターテインメント性を重視した様々な動作を行うことができるため、エンターテインメントロボットと呼称される場合もある。また、そのようなロボット装置には、外部からの情報や内部の状態に応じて自律的に動作するものがある。

【0005】この自律的に動作するロボット装置に用いられる人工知能（AI: artificial intelligence）は、推論・判断等の知的な機能を人工的に実現したものであり、さらに感情や本能等の機能をも人工的に実現することが試みられている。このような人工知能の外部への表現手段としての視覚的な表現手段や聴覚的な表現手段等

7
のうちに、聴覚的なものの一例として、音声を用いることが挙げられる。

【0006】

【発明が解決しようとする課題】ところで、このようなロボット装置に適用する音声合成装置の合成方式としては、録音編集方式や規則音声合成方式等が挙げられる。前者は、人間が発声した単語、文節、定型文等を登録しておき、合成したい言葉に応じて、それらを繋ぎ合わせて文章を作成するものであり、人間の肉声に近い明瞭で自然な音声を得られる可能性があるが、合成できる語彙や文章の種類に限られるという問題がある。一方、後者は、音素や音節などの細かい単位で音声データを蓄積して任意の語彙の合成を可能とするものであるが、音質的に録音編集方式に劣り、特に音素のピッチ、継続時間、音量等の韻律パターンを規則によって付与するため、機械的で不自然な音声となる。

【0007】ここで、両者の欠点を補うため、録音編集方式と規則音声合成方式とを混在させ、定型部分を録音編集で行い、任意語彙部分のみを規則音声合成で行うことも考えられるが、2つの方式で出力音声の音質が異なるため、その繋ぎ目で不自然さが目立つなど、自然で滑らかな音声を合成することができなかった。

【0008】本発明は、このような従来の実情に鑑みて提案されたものであり、感情表現など多様な表情を持った、人間の声に近い自然な音声を合成する音声合成方法、音声合成装置、プログラム及び記録媒体、並びにそのような音声を出力するロボット装置を提供することを目的とする。

【0009】

【課題を解決するための手段】本発明に係る音声合成方法及び装置は、上述した目的を達成するために、入力されたテキストを解析し、当該テキスト中に所定の単語又は文が存在する場合に、予め人間の発声から抽出された上記所定の単語又は文に対応する自然韻律データを記憶手段から選択し、上記所定の単語又は文以外の部分について、上記テキストの音声記号列を作成し、上記音声記号列から韻律データを作成し、上記自然韻律データ又は上記韻律データに基づいて音声を合成することを特徴としている。

【0010】ここで、上記記憶手段には、同一表記に対して、例えば異なる種類の感情又はキャラクタを表す複数種類の自然韻律データが記憶されており、上記自然韻律データ生成する際には、上記複数種類の自然韻律データから、選択制御情報によって指定された1種類が選択される。この選択制御情報は、例えば外部から与えられた情報又は上記テキストに埋め込まれたタグの情報である。

【0011】このような音声合成方法及び装置は、予め記憶手段に種々の単語又は文を記憶しておき、テキストの言語処理の結果、記憶手段に記憶されている単語又は

文については、この記憶手段に登録されている自然韻律データから、例えば外部から与えられた情報又は上記テキストに埋め込まれたタグの情報によって指定された種類の自然韻律データを選択する。一方、登録されていない単語については、通常のテキスト音声合成と同様に、音声記号列を生成してから、韻律データを生成する。そして、韻律データ又は自然韻律データに基づいて音声波形データを生成する。

【0012】また、本発明に係るプログラムは、上述したような音声合成処理をコンピュータに実行させるものであり、本発明に係る記録媒体は、このプログラムが記録されたコンピュータ読み取り可能なものである。

【0013】また、本発明に係るロボット装置は、上述した目的を達成するために、供給された入力情報に基づいて動作を行う自律型のロボット装置であって、所定の単語又は文と、予め人間の発声から抽出された上記所定の単語又は文に対応する自然韻律データとが記憶されている記憶手段と、発話するテキストを解析し、当該テキスト中に上記所定の単語又は文が存在する場合に、当該所定の単語又は文に対応する自然韻律データを上記記憶手段から選択する自然韻律データ選択手段と、上記所定の単語又は文以外の部分について、上記テキストの音声記号列を作成する音声記号列作成手段と、上記音声記号列から韻律データを作成する韻律データ作成手段と、上記自然韻律データ又は上記韻律データに基づいて音声を合成する音声合成手段とを備えることを特徴としている。

【0014】ここで、上記記憶手段には、同一表記に対して、例えば異なる種類の感情又はキャラクタを表す複数種類の自然韻律データが記憶されており、上記自然韻律データ生成手段は、上記複数種類の自然韻律データから、選択制御情報によって指定された1種類を選択する。この選択制御情報は、例えば感情モデルの感情状態情報又は上記テキストに埋め込まれたタグの情報である。

【0015】このようなロボット装置は、予め記憶手段に種々の単語又は文を記憶しておき、発話するテキストの言語処理の結果、記憶手段に記憶されている単語又は文については、この記憶手段に登録されている自然韻律データから、例えば感情モデルの感情状態情報又は上記テキストに埋め込まれたタグの情報によって指定された種類の自然韻律データを選択する。一方、登録されていない単語については、通常のテキスト音声合成と同様に、音声記号列を生成してから、韻律データを生成する。そして、韻律データ又は自然韻律データに基づいて音声波形データを生成する。

【0016】

【発明の実施の形態】以下、本発明を適用した具体的な実施の形態について、図面を参照しながら詳細に説明する。

【0017】先ず、本実施の形態における音声合成装置の概略構成を図1に示す。ここで、この音声合成装置は、少なくとも感情モデル、音声合成手段及び発音手段を有する例えばロボット装置に適用することを想定しているが、これに限定されず、各種ロボット装置や、ロボット以外の各種コンピュータAI (artificial intelligence) 等への適用も可能であることは勿論である。また、以下では、主として日本語の単語或いは文章を合成する場合について説明するが、これに限定されず、種々の言語に適用可能であることは勿論である。

【0018】図1に示すように、音声合成装置200は、言語処理部210と、音声合成部220とで構成される。ここで、言語処理部210は、タグ処理部211と、言語解析部212と、自然韻律辞書記憶部213と、音声記号生成部214とを有している。また、音声合成部220は、韻律生成部221と、韻律データ調整部222と、音素片記憶部223と、波形生成部224とを有している。

【0019】言語処理部210において、タグ処理部211は、入力されたテキストにタグが付されている場合には、通常のテキストとタグとに分割し、テキスト部分を言語解析部212に供給すると共に、タグの情報を言語解析部212に供給する。入力されたテキストにタグが付されていない場合には、タグ処理部211は、入力されたテキストをそのまま言語解析部212に供給する。なお、詳細は後述するが、このタグは、最終的に合成される音声に様々な感情やキャラクタ性等を付与することを指定するものである。

【0020】言語解析部212は、タグ処理部211から供給されたテキストを、図示しない単語辞書記憶部や文法規則記憶部を参照しながら言語処理する。すなわち、単語辞書記憶部には、各単語の品詞情報や、読み、アクセント等の情報が記述された単語辞書が記憶されており、また、文法規則記憶部には、単語辞書記憶部の単語辞書に記述された単語について、単語連鎖に関する制約等の文法規則が記憶されている。そして、言語解析部212は、この単語辞書及び文法規則に基づいて、タグ処理部211から供給されるテキストの形態素解析や構文解析等の解析を行う。ここで、言語解析部212は、自然韻律辞書記憶部213の自然韻律辞書に登録されている単語については、タグ情報を参照しながら、この自然韻律辞書に登録されている自然韻律データを選択し、後述する韻律データ調整部222に供給する。この自然韻律辞書及び自然韻律データについての詳細は後述する。一方、言語処理部212は、自然韻律辞書記憶部213の自然韻律辞書に登録されていない単語については、解析結果を音声記号生成部214に供給する。

【0021】音声記号生成部214は、アクセント規則及びフレーズ規則を参照しながら、言語解析部212から供給される解析結果に基づいて、テキストに対応する

音声記号列を生成する。ここで、アクセント規則とは、アクセントを付与する規則であり、音声記号生成部214は、このアクセント規則に従って、音声記号に、アクセントを表すタグを挿入する。また、フレーズ規則とは、フレーズを決定する際の規則であり、音声記号生成部214は、このフレーズ規則に従って、音声記号に、フレーズを表すタグを挿入する。

【0022】音声合成部220において、韻律生成部221は、音声記号生成部214から供給された音声記号列に基づいて韻律データを作成し、この韻律データを波形合成部224に供給する。この韻律生成部221は、例えば音声記号列から抽出したアクセント型、文中でのアクセント句の数、文中での当該アクセントの位置、当該アクセント句の音韻数、アクセント句内での当該音韻の位置、当該音韻の種類といった情報を用いて、数量化一類等の統計的手法により当該音韻のピッチ周期、継続時間長、音量等のパラメータを表現した韻律データを生成する。

【0023】また、韻律生成部221は、アプリケーションによってピッチ周期、話速、音量等が指定されている場合には、これを考慮して韻律データのパラメータの調整を行う。さらに、韻律生成部221は、タグ情報によって、韻律データのパラメータの調整を行い、感情或いはキャラクタ性の伴われた音声を合成可能とすることができる。

【0024】韻律データ調整部222は、音素片記憶部223から、標準で出力する際の音声の平均ピッチ周期、平均話速、平均音量等のデータを得て、言語解析部212から供給された自然韻律データのパラメータがその平均ピッチ周期等になるように、ピッチ周期、継続時間長、音量の調整を行う。また、韻律データ調整部222は、アプリケーションによってピッチ周期、話速、音量等が指定されている場合には、これを考慮して、自然韻律データのパラメータの調整を行う。

【0025】波形生成部224は、韻律生成部221から供給された韻律データと韻律データ調整部222から供給された自然韻律データとを用いて音声波形を生成する。この波形生成部224は、音素片記憶部223を参照しながら、韻律データ又は自然韻律データに示される音韻系列、ピッチ周期、音量等をもとに、なるべくこれに近い音素片データを検索してその部分を切り出して並べ、音声波形データを生成する。すなわち、音素片記憶部223には、例えば、CV (Consonant, Vowel) や、VCV、CVC等の形で音素片データが記憶されており、波形生成部224は、韻律データ又は自然韻律データに基づいて、必要な音素片データを接続し、さらに、ポーズ、アクセント、イントネーション等を適切に付加することで、音声波形データを生成する。

【0026】得られた音声波形データは、D/A (Digital/Analog) 変換器やアンプ等を介してスピーカに送ら

れることにより、実際の音声として発せられる。例えば、ロボット装置である場合には、このような処理が、いわゆるバーチャルロボットにおいてなされて、スピーカを介して発話されるようになる。

【0027】続いて、以上のような構成を有する音声合成装置200の動作について、図2のフローチャートを用いて説明する。まずステップS1において、発話するためのテキストが入力され、ステップS2において、タグとテキストとが分離される。なお、入力されたテキストにタグが付されていない場合には、ステップS2を省略することができる。

【0028】次にステップS3において、テキストの言語処理が行われる。すなわち、上述したように、各単語の品詞情報や、読み、アクセント等の情報が記述された単語辞書や単語連鎖に関する制約等の文法規則に基づいて、テキストの形態素解析や構文解析等の解析が行われる。

【0029】続いてステップS4において、韻律データ又は自然韻律データが生成される。すなわち、自然韻律辞書に登録されている単語については、自然韻律辞書に登録されている自然韻律データのうち、例えば上述したタグで指定されたものが選択される。また、自然韻律辞書に登録されていない単語については、音声記号列に変換されてから、韻律データが生成される。

【0030】ステップS5では、韻律データ又は自然韻律データのパラメータが調整される。具体的には、自然韻律データは、自然韻律辞書に登録した際のピッチ周期、継続時間長、音量等となっているため、音素片データから標準で出力する際の音声の平均ピッチ周期、平均話速、平均音量等のデータを得て、自然韻律データのパラメータが調整される。

【0031】またステップS5では、タグ情報に基づいて韻律データのパラメータの調整を行い、感情或いはキャラクター性を伴わせることができる。

【0032】さらにステップS5では、アプリケーションによってピッチ周期、話速、音量等が指定されている場合、これを考慮して、韻律データ又は自然韻律データのパラメータが調整される。

【0033】最後にステップS6では、韻律データと自然韻律データとを用いて、音声波形データが生成される。すなわち、韻律データ又は自然韻律データに基づいて、必要な音素片データが接続され、さらに、ポーズ、アクセント、イントネーション等が適切に付加されて、音声波形データが生成される。この音声波形データは、D/A変換器やアンプ等を介してスピーカに送られることにより、実際の音声として発せられる。

【0034】以上説明したように、本実施の形態における音声合成装置200では、予め自然韻律辞書に種々の単語又は文を登録しておき、テキストの言語処理の結果、自然韻律辞書に登録されている単語又は文について

は、この自然韻律辞書に登録されている自然韻律データを選択する。一方、登録されていない単語又は文については、通常のテキスト音声合成と同様に、音声記号列を生成してから、韻律データを生成する。そして、韻律データ又は自然韻律データに基づいて、必要な音素片データを接続し、さらに、ポーズ、アクセント、イントネーション等を適切に付加して、音声波形データを生成する。

【0035】すなわち、種々の単語又は文について人間の自然な発声から得られた自然韻律データを予め自然韻律辞書に登録しておき、登録されていない単語又は文の韻律データと自然韻律データとが、ピッチ周期、継続時間長、音量等のパラメータに基づいて接続される。このように、登録されている単語又は文と登録されていない単語又は文との接続を韻律データのレベルで行うため、より自然性の高い音声を合成することができる。

【0036】なお、上述のように、感情やキャラクタに応じて自然韻律辞書に登録されている単語又は文を選択するような場合、例えばこの自然韻律辞書に予め登録しておく単語としては、感動詞や副詞などの感情が表れやすい単語であることが好ましい。また、これに限定されず、定型的に用いられる単語を自然韻律辞書に登録しておくようにしても構わない。

【0037】以下、具体例を挙げて音声合成装置200の動作についてさらに詳細に説明する。なお、以下では、入力するテキスト例として「¥happiness¥ ねえ、今日は天気が良いね。」を用い、このテキスト中の「ねえ」の部分が自然韻律辞書に登録されているものとして説明する。ここで、¥happiness¥は、そのテキストを喜び(happiness)の感情を伴わせて合成することを意味するタグである。なお、タグがこの例に限定されないことは勿論であり、その他の感情を指定するものであっても構わない。また、感情に限らず、キャラクタを指定するタグが付されていてもよく、さらには、タグが全く付されていない場合でも構わない。

【0038】入力されたタグ付きのテキストは、タグ処理部211(図1)において、タグ(¥happiness¥)とテキスト(「ねえ、今日は天気が良いね。」)とに分離され、このタグの情報とテキストとが言語解析部212に供給される。

【0039】そして、テキスト部分は、言語解析部212において、自然韻律辞書記憶部213の自然韻律辞書を参照しながら言語解析が行われる。ここで、自然韻律辞書は、例えば図3のように構成される。図3に示すように、登録されている単語毎に、標準の自然韻律データのほか、例えば平静(calm)、怒り(anger)、悲しみ(sadness)、喜び(happiness)、落ち着き(comfort)等の各感情や、各キャラクタに対応する自然韻律データが用意される。

【0040】なお、感情の例がこれらに限定されないこ

とは勿論であり、また、各単語について全ての感情に対応する自然韻律データを用意しておかなければならないわけでもない。指定された感情等に対応する自然韻律データが登録されていない場合には、標準の自然韻律データを選択してもよく、また、類似する感情等の自然韻律データを選択するようにしても構わない。例えば、驚きと怖れ、退屈と悲しみなどのある特定の感情に関しては、発せられる音声の音響特性が類似することが知られているため、代替として用いるようにしても構わない。

【0041】本具体例では、テキストにタグ(¥happiness¥)が付されているため、喜び(happiness)に対応する「ねえ」の自然韻律データが選択される。この自然韻律データは、例えば以下の表のように表される。

【0042】

【表1】

[LABEL]	[PITCH]	[VOLUME]
0 .n	0 76	0 100
807 ne	4274 47	807 121
4269 e.	6618 106	4269 101
9596 ..		9596 69

【0043】この表において、[LABEL]は、各音韻の継続時間長を表したものである。すなわち、「.n」という音韻は、0サンプルから807サンプルまでの807サンプルの継続時間長であり、「ne」という音韻は、807サンプルから4269サンプルまでの3462サンプルの継続時間長である。また、[PITCH]は、ピッチ周期を点ピッチで表したものである。すなわち、0サンプルでのピッチ周期は76サンプルであり、4274サンプルでのピッチ周期は47サンプルである。また、[VOLUME]は、各サンプルでの相対的な音量を表したものである。すなわち、デフォルト値を100%としたときに、0サンプルでは100%の音量であり、807サンプルでは121%の音量である。このようにして、全ての音韻が表現される。

【0044】一方、「今日は天気が良いね。」の部分については、自然韻律辞書に登録されていないため、音声記号生成部214に送られ、例えば「Ko' 5ooowa//te' 4xx kiva//yo' 2iine..」というような音声記号列に変換される。ここで、タグ「5」のうちの「'」は、アクセントを表し、続く数字の5は、アクセントの強さを意味する。また、タグ「//」は、アクセント句の区切りを表す。

【0045】このようにして生成された音声記号列は、韻律生成部221において韻律データに変換される。この韻律データは、上述した自然韻律データと同様な構成を有するものであり、各音韻の継続時間長を表した[LABEL]と、ピッチ周期を点ピッチで表した[PITCH]と、各サンプルでの相対的な音量を表した[VOLUME]とで表現される。

【0046】ここで上述したように、入力したテキストにはタグ(¥happiness¥)が付されているため、「ねえ、」の部分と同様に、「今日は天気が良いね。」の部分についても、喜びの感情を表現する必要がある。

【0047】そこで、本実施の形態では、以下の表に示すような、怒り、悲しみ、喜び及び落ち着き等の各感情に対応して予め決定されているパラメータ(少なくとも各音韻の継続時間長(DUR)、ピッチ(PITCH)及び音量(VOLUME)等)の組合せテーブルを各感情の特質に基づいて予め生成しておき、このテーブルを韻律生成部221に保持しておく。ここで、以下のテーブルにおけるピッチの単位はヘルツであり、継続時間長の単位はミリ秒である。

【0048】

【表2】

平静(Calm)	
パラメータ	状態又は値
LASTWORDACCENTED	no
MEANPITCH	280
PITCHVAR	10
MAXPITCH	370
MEANDUR	200
DURVAR	100
PROBACCENT	0.4
DEFAULTCONTORD	rising
CONTOURLASTWORD	rising
VOLUME	100

【0049】

【表3】

怒り(Anger)	
パラメータ	状態又は値
LASTWORDACCENTED	no
MEANPITCH	450
PITCHVAR	100
MAXPITCH	500
MEANDUR	150
DURVAR	20
PROBACCENT	0.4
DEFAULTCONTORD	falling
CONTOURLASTWORD	falling
VOLUME	140

40 【0050】

【表4】

15

悲しみ(sadness)	
パラメータ	状態又は値
LASTWORDACCENTED	nil
MEANPITCH	270
PITCHVAR	30
MAXPITCH	250
MEANDUR	300
DURVAR	100
PROBACCENT	0
DEFAULTCONTORD	falling
CONTOURLASTWORD	falling
VOLUME	90

【0051】

【表5】

落ち着き(comfort)	
パラメータ	状態又は値
LASTWORDACCENTED	t
MEANPITCH	300
PITCHVAR	50
MAXPITCH	350
MEANDUR	300
DURVAR	150
PROBACCENT	0.2
DEFAULTCONTORD	rising
CONTOURLASTWORD	rising
VOLUME	100

【0052】

【表6】

喜び(happiness)	
パラメータ	状態又は値
LASTWORDACCENTED	t
MEANPITCH	400
PITCHVAR	100
MAXPITCH	600
MEANDUR	170
DURVAR	50
PROBACCENT	0.3
DEFAULTCONTORD	rising
CONTOURLASTWORD	rising
VOLUME	120

【0053】このようにして予め用意しておいた各感情に対応されるパラメータからなるテーブルを実際に判別された感情に応じて切り換え、このテーブルに基づいてパラメータを変更することにより、感情を表現することが可能とされる。

【0054】具体的には、欧州特許出願第01401880.1号の明細書及び図面に記載された技術を応用することができる。

【0055】例えば発話される単語に含まれる音韻の平均ピッチ周期がMEANPITCHの値に基づいて計算される値になるように各音韻のピッチ周期を変化させ、また、ピッチ周期の分散値がPITCHVARの値に基づいて計算される値になるように制御する。

16

【0056】同様に、発話される単語に含まれる音韻の平均継続時間長がMEANDURの値で計算される値になるように各音韻の継続時間長を変化させ、また、継続時間長の分散値がDURVARの値になるように制御する。

【0057】また、各音韻の音量についても、各感情のテーブルにおけるVOLUMEで指定される値に制御する。

【0058】さらに、このテーブルに基づいて各アクセント句のcontour（輪郭）を変更することもできる。すなわち、DEFAULTCONTOUR=risingである場合には、アクセント句のピッチの傾きが上り調子となり、DEFAULTCONTOUR=fallingである場合には、反対に下り調子となる。

【0059】このように感情に応じて選択されたテーブルを使用してパラメータの変更された韻律データは、波形生成部224に送られる。

【0060】なお、アプリケーションによってピッチ周期、話速、音量等が設定されている場合には、このデータによっても韻律データのピッチ周期、継続時間長、音量といったパラメータの調整が行われる。

20 【0061】一方、「ねえ、」の部分の自然韻律データは、韻律データ調整部222において、ピッチ周期、継続時間長、音量といったパラメータの調整が行われる。すなわち、自然韻律データは、自然韻律辞書に登録した際のピッチ周期、継続時間長、音量等となっているため、波形生成部224が用いる音素片データから、標準で出力する際の音声の平均ピッチ周期、平均話速、平均音量等のデータを得て、自然韻律データのパラメータの調整が行われる。

30 【0062】また、韻律データの平均ピッチ周期が上述したように喜びの感情に対応したテーブルの平均ピッチ周期となるように変更されているため、自然韻律データについても、このテーブルの平均ピッチ周期となるように調整される。

【0063】さらに、アプリケーションによってピッチ周期、話速、音量等が設定されている場合には、このデータによっても自然韻律データのパラメータの調整が行われる。

【0064】このようにしてパラメータの調整された自然韻律データは、波形生成部224に送られる。

40 【0065】そして、上述したように、波形生成部224において、この韻律データと自然韻律データとを用いて、音声波形データが生成される。すなわち、韻律データ又は自然韻律データに基づいて、必要な音素片データが接続され、さらに、ポーズ、アクセント、イントネーション等が適切に付加されて、音声波形データが生成される。この音声波形データは、D/A変換器やアンプ等を介してスピーカに送られることにより、実際の音声として発せられる。

50 【0066】なお、以上の説明では、合成音声に対してタグで指定された感情又はキャラクター性を伴わせる例に

ついて説明したが、これに限定されるものではなく、外部から与えられた感情状態情報やキャラクタ情報によって指定された感情又はキャラクタ性を合成音声に伴わせるようにしても構わない。

【0067】感情を例に挙げれば、例えばロボット装置の場合には、行動モデルとして、内部に確率状態遷移モデル（例えば、後述するように、状態遷移表を有するモデル）を有しており、各状態が認識結果や感情や本能の値によって異なる遷移確率テーブルを持ち、その確率に従って次の状態へ遷移し、この遷移に関連付けられた行動を出力する。

【0068】感情による喜びや悲しみの表現行動がこの確率状態遷移モデル（或いは確率遷移表）に記述されており、この表現行動の1つとして、音声による（発話による）感情表現が含まれている。

【0069】すなわち、このロボット装置では、感情モデルの感情状態を表すパラメータを行動モデルが参照することにより決定される行動の1つの要素として感情表現があり、行動決定部の一部機能として、感情状態の判別が行われることになる。そして、この判別された感情状態情報が上述した言語解析部212及び韻律生成部221に与えられる。これにより、その感情に応じた自然韻律データが選択され、また、その感情に応じて韻律データ及び自然韻律データのパラメータが調整される。

【0070】以下、このようなロボット装置の一例として、2本足の自律型ロボットに本発明を適用した例について、図面を参照しながら詳細に説明する。この人間型ロボット装置のソフトウェアに感情・本能モデルを導入し、より人間に近い行動を得ることができるようにしている。本実施の形態では実際に動作をするロボットを用いているが、発話はスピーカを持つコンピュータ・システムであれば容易に実現可能であり、人間と機械とのインタラクション（或いは対話）の場で有効な機能である。したがって本発明の適用範囲はロボットシステムに限られるものではない。

【0071】具体例として図4に示す人間型のロボット装置は、住環境その他の日常生活上の様々な場面における人的活動を支援する実用ロボットであり、内部状態（怒り、悲しみ、喜び、楽しみ等）に応じて行動できるほか、人間が行う基本的な動作を表出できるエンターテインメントロボットである。

【0072】図4に示すように、ロボット装置1は、体幹部ユニット2の所定の位置に頭部ユニット3が連結されると共に、左右2つの腕部ユニット4R/Lと、左右2つの脚部ユニット5R/Lが連結されて構成されている（但し、R及びLの各々は、右及び左の各々を示す接尾辞である。以下において同じ。）。)

【0073】このロボット装置1が具備する関節自由度構成を図5に模式的に示す。頭部ユニット3を支持する首関節は、首関節ヨー軸101と、首関節ピッチ軸10

2と、首関節ロール軸103という3自由度を有している。

【0074】また、上肢を構成する各々の腕部ユニット4R/Lは、肩関節ピッチ軸107と、肩関節ロール軸108と、上腕ヨー軸109と、肘関節ピッチ軸110と、前腕ヨー軸111と、手首関節ピッチ軸112と、手首関節ロール軸113と、手部114とで構成される。手部114は、実際には、複数本の指を含む多関節・多自由度構造体である。ただし、手部114の動作は、ロボット装置1の姿勢制御や歩行制御に対する寄与や影響が少ないので、本明細書ではゼロ自由度と仮定する。したがって、各腕部は7自由度を有するとする。

【0075】また、体幹部ユニット2は、体幹ピッチ軸104と、体幹ロール軸105と、体幹ヨー軸106という3自由度を有する。

【0076】また、下肢を構成する各々の脚部ユニット5R/Lは、股関節ヨー軸115と、股関節ピッチ軸116と、股関節ロール軸117と、膝関節ピッチ軸118と、足首関節ピッチ軸119と、足首関節ロール軸120と、足部121とで構成される。本明細書中では、股関節ピッチ軸116と股関節ロール軸117の交点は、ロボット装置1の股関節位置を定義する。人体の足部121は、実際には多関節・多自由度の足底を含んだ構造体であるが、ロボット装置1の足底は、ゼロ自由度とする。したがって、各脚部は、6自由度で構成される。

【0077】以上を総括すれば、ロボット装置1全体としては、合計で $3+7 \times 2+3+6 \times 2=32$ 自由度を有することになる。ただし、エンターテインメント向けのロボット装置1が必ずしも32自由度に限定されるわけではない。設計・制作上の制約条件や要求仕様等に応じて、自由度すなわち関節数を適宜増減することができるというまでもない。

【0078】上述したようなロボット装置1が持つ各自由度は、実際にはアクチュエータを用いて実装される。外観上で余分な膨らみを排してヒトの自然体形状に近似させること、2足歩行という不安定構造体に対して姿勢制御を行うことなどの要請から、アクチュエータは小型且つ軽量であることが好ましい。

【0079】ロボット装置1の制御システム構成を図6に模式的に示す。図6に示すように、体幹部ユニット2には、CPU (Central Processing Unit) 10、DRAM (Dynamic Random Access Memory) 11、フラッシュROM (Read Only Memory) 12、PC (Personal Computer) カードインターフェース回路13及び信号処理回路14が内部バス15を介して相互に接続されることにより形成されたコントロール部16と、このロボット装置1の動力源としてのバッテリー17とが収納されている。また、体幹部ユニット2には、ロボット装置1の向きや動きの加速度を検出するための角速度センサ18

及び加速度センサ19なども収納されている。

【0080】また、頭部ユニット3には、外部の状況を撮像するための左右の「眼」に相当するCCD (Charge Coupled Device) カメラ20R/Lと、そのCCDカメラ20R/Lからの画像データに基づいてステレオ画像データを作成するための画像処理回路21と、使用者からの「撫でる」や「叩く」といった物理的な働きかけにより受けた圧力を検出するためのタッチセンサ22と、各脚部ユニット5R/Lの足底が着床したか否かを検出する接地確認センサ23R/Lと、姿勢を計測する姿勢センサ24と、前方に位置する物体までの距離を測定するための距離センサ25と、外部音を集音するためのマイクロホン26と、スピーチ等の音声を出力するためのスピーカ27と、LED (Light Emitting Diode) 28などがそれぞれ所定位置に配置されている。

【0081】ここで、接地確認センサ23R/Lは、例えば足底に設置された近接センサ又はマイクロ・スイッチなどで構成される。また、姿勢センサ24は、例えば、加速度センサとジャイロ・センサの組み合わせによって構成される。接地確認センサ23R/Lの出力によって、歩行・走行などの動作期間中において、左右の各脚部ユニット5R/Lが現在立脚又は遊脚何れの状態であるかを判別することができる。また、姿勢センサ24の出力により、体幹部分の傾きや姿勢を検出することができる。

【0082】さらに、体幹部ユニット2、腕部ユニット4R/L、脚部ユニット5R/Lの各関節部分などにはそれぞれ上述した自由度数分のアクチュエータ29₁～29_n及びポテンショメータ30₁～30_nが配設されている。例えば、アクチュエータ29₁～29_nはサーボモータを構成として有している。サーボモータの駆動により、例えば腕部ユニット4R/L及び脚部ユニット5R/Lが制御されて、目標の姿勢或いは動作に遷移する。

【0083】そして、これら角速度センサ18、加速度センサ19、タッチセンサ22、接地確認センサ23R/L、姿勢センサ24、距離センサ25、マイクロホン26、スピーカ27及び各ポテンショメータ30₁～30_nなどの各種センサ並びにLED28及び各アクチュエータ29₁～29_nは、それぞれ対応するハブ31₁～31_nを介してコントロール部16の信号処理回路14と接続され、バッテリー17及び画像処理回路21は、それぞれ信号処理回路14と直接接続されている。

【0084】信号処理回路14は、上述の各センサから供給されるセンサデータや画像データ及び音声データを順次取り込み、これらをそれぞれ内部バス15を介してDRAM11内の所定位置に順次格納する。また信号処理回路14は、これと共にバッテリー17から供給されるバッテリー残量を表すバッテリー残量データを順次取り込み、これをDRAM11内の所定位置に格納する。

【0085】このようにしてDRAM11に格納された各センサデータ、画像データ、音声データ及びバッテリー残量データは、この後CPU10がこのロボット装置1の動作制御を行う際に利用される。

【0086】實際上CPU10は、ロボット装置1の電源が投入された初期時、体幹部ユニット2の図示しないPCカードスロットに装填されたメモリカード32又はフラッシュROM12に格納された制御プログラムをPCカードインターフェース回路13を介して又は直接読み出し、これをDRAM11に格納する。

【0087】また、CPU10は、この後上述のように信号処理回路14よりDRAM11に順次格納される各センサデータ、画像データ、音声データ及びバッテリー残量データに基づいて自己及び周囲の状況や、使用者からの指示及び働きかけの有無などを判断する。

【0088】さらに、CPU10は、この判断結果及びDRAM11に格納した制御プログラムに基づいて続く行動を決定すると共に、当該決定結果に基づいて必要なアクチュエータ29₁～29_nを駆動させることにより、各腕部ユニット4R/Lを上下左右に振らせたり、各脚部ユニット5R/Lを駆動させて歩行させるなどの行動を行わせる。

【0089】また、この際CPU10は、必要に応じて音声データを生成し、これを信号処理回路14を介して音声信号としてスピーカ27に与えることにより当該音声信号に基づく音声を外部に出力させたり、上述のLED28を点灯、消灯又は点滅させる。

【0090】このようにしてこのロボット装置1においては、自己及び周囲の状況や、使用者からの指示及び働きかけに応じて自律的に行動し得るようになされている。

【0091】ところで、このロボット装置1は、内部状態に応じて自律的に行動することができる。そこで、ロボット装置1における制御プログラムのソフトウェア構成例について、図7乃至図12を用いて説明する。なお、この制御プログラムは、上述したように、予めフラッシュROM12に格納されており、ロボット装置1の電源投入初期時において読み出される。

【0092】図7において、デバイス・ドライバ・レイヤ40は、制御プログラムの最下位層に位置し、複数のデバイス・ドライバからなるデバイス・ドライバ・セット41から構成されている。この場合、各デバイス・ドライバは、CCDカメラやタイマ等の通常のコンピュータで用いられるハードウェアに直接アクセスすることを許されたオブジェクトであり、対応するハードウェアからの割り込みを受けて処理を行う。

【0093】また、ロボティック・サーバ・オブジェクト42は、デバイス・ドライバ・レイヤ40の最下位層に位置し、例えば上述の各種センサやアクチュエータ29₁～29_n等のハードウェアにアクセスするためのイ

ンターフェースを提供するソフトウェア群でなるバーチャル・ロボット43と、電源の切換えなどを管理するソフトウェア群でなるパワーマネージャ44と、他の種々のデバイス・ドライバを管理するソフトウェア群でなるデバイス・ドライバ・マネージャ45と、ロボット装置1の機構を管理するソフトウェア群でなるデザインド・ロボット46とから構成されている。

【0094】マネージャ・オブジェクト47は、オブジェクト・マネージャ48及びサービス・マネージャ49から構成されている。オブジェクト・マネージャ48は、ロボティック・サーバ・オブジェクト42、ミドル・ウェア・レイヤ50、及びアプリケーション・レイヤ51に含まれる各ソフトウェア群の起動や終了を管理するソフトウェア群であり、サービス・マネージャ49は、メモリカードに格納されたコネクションファイルに記述されている各オブジェクト間の接続情報に基づいて各オブジェクトの接続を管理するソフトウェア群である。

【0095】ミドル・ウェア・レイヤ50は、ロボティック・サーバ・オブジェクト42の上位層に位置し、画像処理や音声処理などのこのロボット装置1の基本的な機能を提供するソフトウェア群から構成されている。また、アプリケーション・レイヤ51は、ミドル・ウェア・レイヤ50の上位層に位置し、当該ミドル・ウェア・レイヤ50を構成する各ソフトウェア群によって処理された処理結果に基づいてロボット装置1の行動を決定するためのソフトウェア群から構成されている。

【0096】なお、ミドル・ウェア・レイヤ50及びアプリケーション・レイヤ51の具体的なソフトウェア構成をそれぞれ図8に示す。

【0097】ミドル・ウェア・レイヤ50は、図8に示すように、騒音検出用、温度検出用、明るさ検出用、音階認識用、距離検出用、姿勢検出用、タッチセンサ用、動き検出用及び色認識用の各信号処理モジュール60～68並びに入力セマンティクスコンバータモジュール69などを有する認識系70と、出力セマンティクスコンバータモジュール78並びに姿勢管理用、トラッキング用、モーション再生用、歩行用、転倒復帰用、LED点灯用及び音再生用の各信号処理モジュール71～77などを有する出力系79とから構成されている。

【0098】認識系70の各信号処理モジュール60～68は、ロボティック・サーバ・オブジェクト42のバーチャル・ロボット43によりDRAMから読み出される各センサデータや画像データ及び音声データのうちの対応するデータを取り込み、当該データに基づいて所定の処理を施して、処理結果を入力セマンティクスコンバータモジュール69に与える。ここで、例えば、バーチャル・ロボット43は、所定の通信規約によって、信号の授受或いは変換をする部分として構成されている。

【0099】入力セマンティクスコンバータモジュール

69は、これら各信号処理モジュール60～68から与えられる処理結果に基づいて、「うるさい」、「暑い」、「明るい」、「ボールを検出した」、「転倒を検出した」、「撫でられた」、「叩かれた」、「ドミソの音階が聞こえた」、「動く物体を検出した」又は「障害物を検出した」などの自己及び周囲の状況や、使用者からの指令及び働きかけを認識し、認識結果をアプリケーション・レイヤ41に出力する。

【0100】アプリケーション・レイヤ51は、図9に示すように、行動モデルライブラリ80、行動切換モジュール81、学習モジュール82、感情モデル83及び本能モデル84の5つのモジュールから構成されている。

【0101】行動モデルライブラリ80には、図10に示すように、「バッテリー残量が少なくなった場合」、「転倒復帰する」、「障害物を回避する場合」、「感情を表現する場合」、「ボールを検出した場合」などの予め選択されたいくつかの条件項目にそれぞれ対応させて、それぞれ独立した行動モデルが設けられている。

【0102】そして、これら行動モデルは、それぞれ入力セマンティクスコンバータモジュール69から認識結果が与えられたときや、最後の認識結果が与えられてから一定時間が経過したときなどに、必要に応じて後述のように感情モデル83に保持されている対応する情動のパラメータ値や、本能モデル84に保持されている対応する欲求のパラメータ値を参照しながら続く行動をそれぞれ決定し、決定結果を行動切換モジュール81に出力する。

【0103】なお、この実施の形態の場合、各行動モデルは、次の行動を決定する手法として、図11に示すような1つのノード(状態) $NODE_0 \sim NODE_n$ から他のどのノード $NODE_0 \sim NODE_n$ に遷移するかを各ノード $NODE_0 \sim NODE_n$ に間を接続するアーク $ARC_1 \sim ARC_n$ に対してそれぞれ設定された遷移確率 $P_1 \sim P_n$ に基づいて確率的に決定する有限確率オートマトンと呼ばれるアルゴリズムを用いる。

【0104】具体的に、各行動モデルは、それぞれ自己の行動モデルを形成するノード $NODE_0 \sim NODE_n$ にそれぞれ対応させて、これらノード $NODE_0 \sim NODE_n$ 毎に図12に示すような状態遷移表90を有している。

【0105】この状態遷移表90では、そのノード $NODE_0 \sim NODE_n$ において遷移条件とする入力イベント(認識結果)が「入力イベント名」の列に優先順に列記され、その遷移条件についてのさらなる条件が「データ名」及び「データ範囲」の列における対応する行に記述されている。

【0106】したがって、図12の状態遷移表90で表されるノード $NODE_{100}$ では、「ボールを検出(BALL)」という認識結果が与えられた場合に、当該認

識結果とともに与えられるそのボールの「大きさ (SIZE)」が「0から1000」の範囲であることや、「障害物を検出 (OBSTACLE)」という認識結果が与えられた場合に、当該認識結果とともに与えられるその障害物までの「距離 (DISTANCE)」が「0から100」の範囲であることが他のノードに遷移するための条件となっている。

【0107】また、このノード $NODE_{100}$ では、認識結果の入力がない場合においても、行動モデルが周期的に参照する感情モデル83及び本能モデル84にそれぞれ保持された各情動及び各欲求のパラメータ値のうち、感情モデル83に保持された「喜び (Joy)」、「驚き (Surprise)」又は「悲しみ (Sadness)」の何れかのパラメータ値が「50から100」の範囲であるときには他のノードに遷移することができるようになってい

る。

【0108】また、状態遷移表90では、「他のノードへの遷移確率」の欄における「遷移先ノード」の行にそのノード $NODE_0 \sim NODE_n$ から遷移できるノード名が列記されているとともに、「入力イベント名」、「データ名」及び「データの範囲」の列に記述された全ての条件が揃ったときに遷移できる他の各ノード $NODE_0 \sim NODE_n$ への遷移確率が「他のノードへの遷移確率」の欄内の対応する箇所にそれぞれ記述され、そのノード $NODE_0 \sim NODE_n$ に遷移する際に出力すべき行動が「他のノードへの遷移確率」の欄における「出力行動」の行に記述されている。なお、「他のノードへの遷移確率」の欄における各行の確率の和は100 [%]となっている。

【0109】したがって、図12の状態遷移表90で表されるノード $NODE_{100}$ では、例えば「ボールを検出 (BALL)」し、そのボールの「SIZE (大きさ)」が「0から1000」の範囲であるという認識結果が与えられた場合には、「30 [%]」の確率で「ノード $NODE_{120}$ (node 120)」に遷移でき、そのとき「ACTION1」の行動が出力されることとなる。

【0110】各行動モデルは、それぞれこのような状態遷移表90として記述されたノード $NODE_0 \sim NODE_n$ が幾つも繋がるようにして構成されており、入力セマンティクスコンバータモジュール69から認識結果が与えられたときなどに、対応するノード $NODE_0 \sim NODE_n$ の状態遷移表を利用して確率的に次の行動を決定し、決定結果を行動切換モジュール81に出力するようになされている。

【0111】図9に示す行動切換モジュール81は、行動モデルライブラリ80の各行動モデルからそれぞれ出力される行動のうち、予め定められた優先順位の高い行動モデルから出力された行動を選択し、当該行動を実行

すべき旨のコマンド（以下、行動コマンドという。）をミドル・ウェア・レイヤ50の出力セマンティクスコンバータモジュール78に送出する。なお、この実施の形態においては、図10において下側に表記された行動モデルほど優先順位が高く設定されている。

【0112】また、行動切換モジュール81は、行動完了後に出力セマンティクスコンバータモジュール78から与えられる行動完了情報に基づいて、その行動が完了したことを学習モジュール82、感情モデル83及び本能モデル84に通知する。

【0113】一方、学習モジュール82は、入力セマンティクスコンバータモジュール69から与えられる認識結果のうち、「叩かれた」や「撫でられた」など、使用者からの働きかけとして受けた教示の認識結果を入力する。

【0114】そして、学習モジュール82は、この認識結果及び行動切換モジュール71からの通知に基づいて、「叩かれた（叱られた）」ときにはその行動の発現確率を低下させ、「撫でられた（誉められた）」ときにはその行動の発現確率を上昇させるように、行動モデルライブラリ70における対応する行動モデルの対応する遷移確率を変更する。

【0115】他方、感情モデル83は、「喜び (Joy)」、「悲しみ (Sadness)」、「怒り (Anger)」、「驚き (Surprise)」、「嫌悪 (Disgust)」及び「恐れ (Fear)」の合計6つの情動について、各情動毎にその情動の強さを表すパラメータを保持している。そして、感情モデル83は、これら各情動のパラメータ値を、それぞれ入力セマンティクスコンバータモジュール69から与えられる「叩かれた」及び「撫でられた」などの特定の認識結果や、経過時間及び行動切換モジュール81からの通知などに基づいて周期的に更新する。

【0116】具体的には、感情モデル83は、入力セマンティクスコンバータモジュール69から与えられる認識結果と、そのときのロボット装置1の行動と、前回更新してから経過時間などに基づいて所定の演算式により算出されるそのときのその情動の変動量を ΔE [t]、現在のその情動のパラメータ値を E [t]、その情動の感度を表す係数を k 。として、(1)式によって次の周期におけるその情動のパラメータ値 E [t+1]を算出し、これを現在のその情動のパラメータ値 E [t]と置き換えるようにしてその情動のパラメータ値を更新する。また、感情モデル83は、これと同様にして全ての情動のパラメータ値を更新する。

【0117】

【数1】

$$E[t+1] = E[t] + k_e \times \Delta E[t]$$

【0118】なお、各認識結果や出力セマンティクスコンバータモジュール78からの通知が各情動のパラメータ値の変動量 $\Delta E[t]$ にどの程度の影響を与えるかは予め決められており、例えば「叩かれた」といった認識結果は「怒り」の情動のパラメータ値の変動量 $\Delta E[t]$ に大きな影響を与え、「撫でられた」といった認識結果は「喜び」の情動のパラメータ値の変動量 $\Delta E[t]$ に大きな影響を与えるようになっている。

【0119】ここで、出力セマンティクスコンバータモジュール78からの通知とは、いわゆる行動のフィードバック情報（行動完了情報）であり、行動の出現結果の情報であり、感情モデル83は、このような情報によっても感情を変化させる。これは、例えば、「叫ぶ」といった行動により怒りの感情レベルが下がるといったようなことである。なお、出力セマンティクスコンバータモジュール78からの通知は、上述した学習モジュール82にも入力されており、学習モジュール82は、その通知に基づいて行動モデルの対応する遷移確率を変更する。

【0120】なお、行動結果のフィードバックは、行動切換モジュール81の出力（感情が付加された行動）によりなされるものであってもよい。

【0121】一方、本能モデル84は、「運動欲（exer*

... (1)

「cise）」、「愛情欲（affection）」、「食欲（appetite）」及び「好奇心（curiosity）」の互いに独立した4つの欲求について、これら欲求毎にその欲求の強さを表すパラメータを保持している。そして、本能モデル84は、これらの欲求のパラメータ値を、それぞれ入力セマンティクスコンバータモジュール69から与えられる認識結果や、経過時間及び行動切換モジュール81からの通知などに基づいて周期的に更新する。

【0122】具体的には、本能モデル84は、「運動欲」、「愛情欲」及び「好奇心」については、認識結果、経過時間及び出力セマンティクスコンバータモジュール78からの通知などに基づいて所定の演算式により算出されるそのときのその欲求の変動量を $\Delta I[k]$ 、現在のその欲求のパラメータ値を $I[k]$ 、その欲求の感度を表す係数 k_i として、所定周期で(2)式を用いて次の周期におけるその欲求のパラメータ値 $I[k+1]$ を算出し、この演算結果を現在のその欲求のパラメータ値 $I[k]$ と置き換えるようにしてその欲求のパラメータ値を更新する。また、本能モデル84は、これと同様に「食欲」を除く各欲求のパラメータ値を更新する。

【0123】

【数2】

... (2)

$$I[k+1] = I[k] + k_i \times \Delta I[k]$$

【0124】なお、認識結果及び出力セマンティクスコンバータモジュール78からの通知などが各欲求のパラメータ値の変動量 $\Delta I[k]$ にどの程度の影響を与えるかは予め決められており、例えば出力セマンティクスコンバータモジュール78からの通知は、「疲れ」のパラメータ値の変動量 $\Delta I[k]$ に大きな影響を与えるようになっている。

【0125】なお、本実施の形態においては、各情動及び各欲求（本能）のパラメータ値がそれぞれ0から100までの範囲で変動するように規制されており、また係数 k_e 、 k_i の値も各情動及び各欲求毎に個別に設定されている。

【0126】一方、ミドル・ウェア・レイヤ50の出力セマンティクスコンバータモジュール78は、図8に示すように、上述のようにしてアプリケーション・レイヤ51の行動切換モジュール81から与えられる「前進」、「喜ぶ」、「話す」又は「トラッキング（ボールを追いかける）」といった抽象的な行動コマンドを出力系79の対応する信号処理モジュール71～77に与える。

【0127】そしてこれら信号処理モジュール71～77は、行動コマンドが与えられると当該行動コマンドに基づいて、その行動をするために対応するアクチュエータに与えるべきサーボ指令値や、スピーカから出力する

音の音声データ及び又はLEDに与える駆動データを生成し、これらのデータをロボティック・サーバ・オブジェクト42のバーチャル・ロボット43及び信号処理回路を順次介して対応するアクチュエータ又はスピーカ又はLEDに順次送出する。

【0128】このようにしてロボット装置1は、上述した制御プログラムに基づいて、自己（内部）及び周囲（外部）の状況や、使用者からの指示及び働きかけに応じた自律的な行動ができる。

【0129】このような制御プログラムは、ロボット装置が読取可能な形式で記録された記録媒体を介して提供される。制御プログラムを記録する記録媒体としては、磁気読取方式の記録媒体（例えば、磁気テープ、フレキシブルディスク、磁気カード）、光学読取方式の記録媒体（例えば、CD-ROM、MO、CD-R、DVD）等が考えられる。記録媒体には、半導体メモリ（いわゆるメモリカード（矩形型、正方形型など形状は問わない。）、ICカード）等の記憶媒体も含まれる。また、制御プログラムは、いわゆるインターネット等を介して提供されてもよい。

【0130】これらの制御プログラムは、専用の読込ドライブ装置、又はパーソナルコンピュータ等を介して再生され、有線又は無線接続によってロボット装置1に伝送されて読み込まれる。また、ロボット装置1は、半導体メモリ、又はICカード等の小型化された記憶媒体のドライブ装置を備える場合、これら記憶媒体から制御プログラムを直接読み込むこともできる。

【0131】以上のように構成されたロボット装置1において、上述した音声合成のアルゴリズムは、図8中の音再生モジュール77として実装される。音再生モジュール77では、上位の部分（例えば、行動モデル）にて決定された音出力コマンド（例えば、「喜びで発話せよ」など）を受け、実際の音声波形データを生成し、順にバーチャルロボット43のスピーカデバイスに対してデータを送信する。これによりロボット装置において、図6に示すスピーカ27から感情表現がなされた発話文が発せられる。

【0132】なお、本発明は上述した実施の形態のみに限定されるものではなく、本発明の要旨を逸脱しない範囲において種々の変更が可能であることは勿論である。

【0133】

【発明の効果】以上詳細に説明したように、本発明に係る音声合成方法及び装置は、入力されたテキストを解析し、当該テキスト中に所定の単語又は文が存在する場合に、予め人間の発声から抽出された上記所定の単語又は文に対応する自然韻律データを記憶手段から選択し、上記所定の単語又は文以外の部分について、上記テキストの音声記号列を作成し、上記音声記号列から韻律データを作成し、上記自然韻律データ又は上記韻律データに基づいて音声を作成することを特徴としている。

【0134】ここで、上記記憶手段には、同一表記に対して、例えば異なる種類の感情又はキャラクタを表す複数種類の自然韻律データが記憶されており、上記自然韻律データ生成する際には、上記複数種類の自然韻律データから、選択制御情報によって指定された1種類が選択される。この選択制御情報は、例えば外部から与えられた情報又は上記テキストに埋め込まれたタグの情報である。

【0135】このような音声合成方法及び装置は、予め記憶手段に種々の単語又は文を記憶しておき、テキストの言語処理の結果、記憶手段に記憶されている単語又は文については、この記憶手段に登録されている自然韻律データから、例えば外部から与えられた情報又は上記テキストに埋め込まれたタグの情報によって指定された種類の自然韻律データを選択する。一方、登録されていない単語又は文については、通常のテキスト音声合成と同様に、音声記号列を生成してから、韻律データを生成する。そして、韻律データ又は自然韻律データに基づいて音声波形データを生成する。これにより、より人間の声に近い自然な音声を合成することが可能となる。

【0136】また、本発明に係るプログラムは、上述したような音声合成処理をコンピュータに実行させるものであり、本発明に係る記録媒体は、このプログラムが記録されたコンピュータ読み取り可能なものである。

【0137】このようなプログラム及び記録媒体によれば、登録されている単語又は文の自然韻律データと登録されていない単語又は文の韻律データとを合成することにより、より人間の声に近い自然な音声を合成することが可能となる。

【0138】また、本発明に係るロボット装置は、供給された入力情報に基づいて動作を行う自律型のロボット装置であって、所定の単語又は文と、予め人間の発声から抽出された上記所定の単語又は文に対応する自然韻律データとが記憶されている記憶手段と、発話するテキストを解析し、当該テキスト中に上記所定の単語又は文が存在する場合に、当該所定の単語又は文に対応する自然韻律データを上記記憶手段から選択する自然韻律データ選択手段と、上記所定の単語又は文以外の部分について、上記テキストの音声記号列を作成する音声記号列作成手段と、上記音声記号列から韻律データを作成する韻律データ作成手段と、上記自然韻律データ又は上記韻律データに基づいて音声を合成する音声合成手段とを備えることを特徴としている。

【0139】ここで、上記記憶手段には、同一表記に対して、例えば異なる種類の感情又はキャラクタを表す複数種類の自然韻律データが記憶されており、上記自然韻律データ生成手段は、上記複数種類の自然韻律データから、選択制御情報によって指定された1種類を選択する。この選択制御情報は、例えば感情モデルの感情状態情報又は上記テキストに埋め込まれたタグの情報であ

る。

【0140】このようなロボット装置は、予め記憶手段に種々の単語又は文を記憶しておき、発話するテキストの言語処理の結果、記憶手段に記憶されている単語又は文については、この記憶手段に登録されている自然韻律データから、例えば感情モデルの感情状態情報又は上記テキストに埋め込まれたタグの情報によって指定された種類の自然韻律データを選択する。一方、登録されていない単語については、通常のテキスト音声合成と同様に、音声記号列を生成してから、韻律データを生成する。そして、韻律データ又は自然韻律データに基づいて音声波形データを生成する。これにより、より人間の声に近い自然な音声を合成することが可能となる。

【図面の簡単な説明】

【図1】本実施の形態における音声合成装置の概略構成を説明する図である。

【図2】同音声合成装置の動作を説明するフローチャートである。

【図3】同音声合成装置における自然韻律辞書の構成例を説明する図である。

【図4】本実施の形態におけるロボット装置の外観構成を示す斜視図である。

【図5】同ロボット装置の自由度構成モデルを模式的に示す図である。

*【図6】同ロボット装置の回路構成を示すブロック図である。

【図7】同ロボット装置のソフトウェア構成を示すブロック図である。

【図8】同ロボット装置のソフトウェア構成におけるミドル・ウェア・レイヤの構成を示すブロック図である。

【図9】同ロボット装置のソフトウェア構成におけるアプリケーション・レイヤの構成を示すブロック図である。

10 【図10】アプリケーション・レイヤの行動モデルライブラリの構成を示すブロック図である。

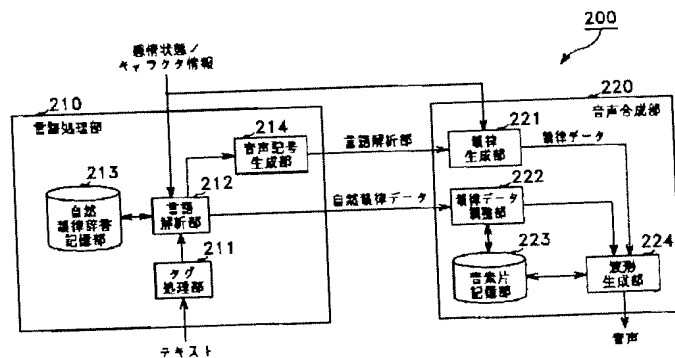
【図11】同ロボット装置の行動決定のための情報となる有限確率オートマトンを説明する図である。

【図12】有限確率オートマトンの各ノードに用意された状態遷移表を示す図である。

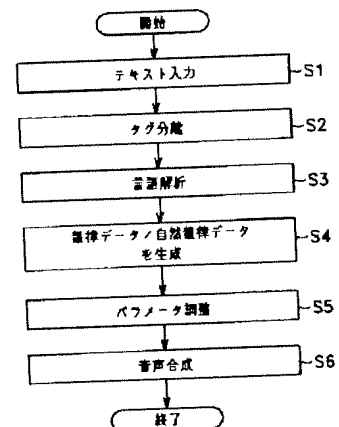
【符号の説明】

1 ロボット装置、10 CPU、14 信号処理回路、27 スピーカ、80 行動モデル、83 感情モデル、200 音声合成装置、210 言語処理部、211 タグ処理部、212 言語解析部、213 自然韻律辞書記憶部、214 音声記号生成部、220 音声合成部、221 韻律生成部、222 音素データ記憶部、223 音素片記憶部、224 波形生成部

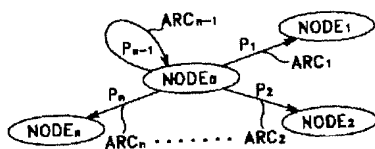
【図1】



【図2】



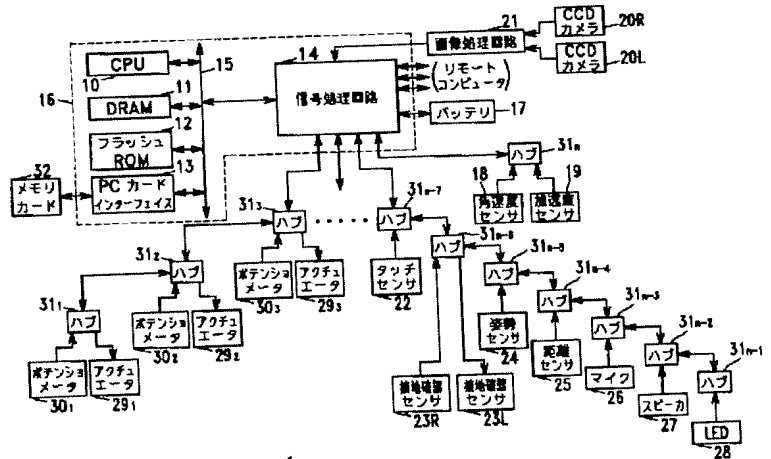
【図11】



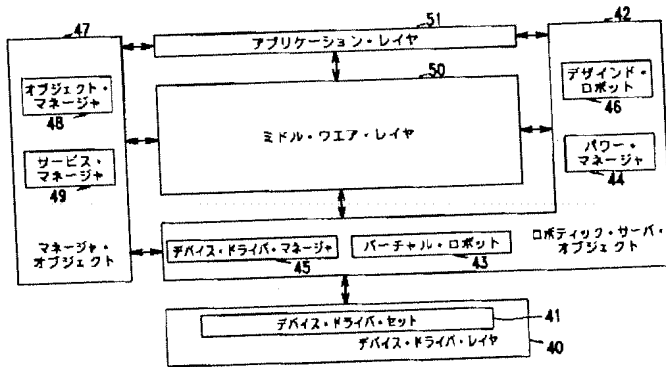
【図 4】

Figure 1 is a perspective view of a humanoid robot 1. The robot has a head 3, a torso 2, and two arms 4R and 4L. It also has two legs 5R and 5L. The robot is standing on a base.

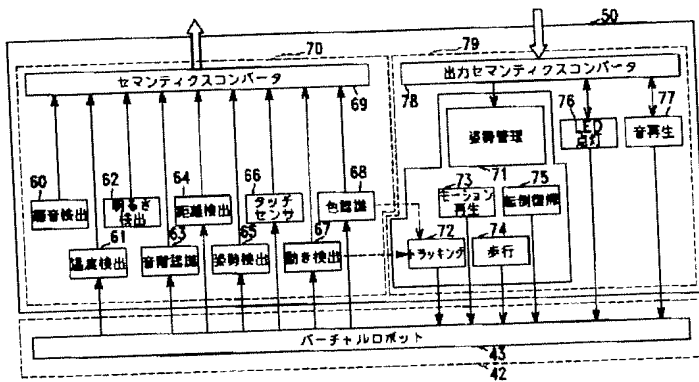
【图 6】



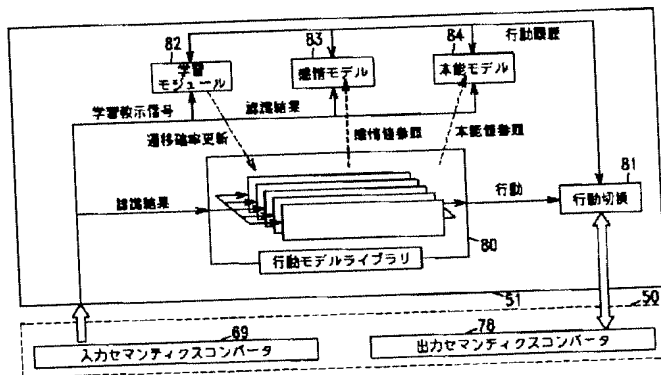
【図7】



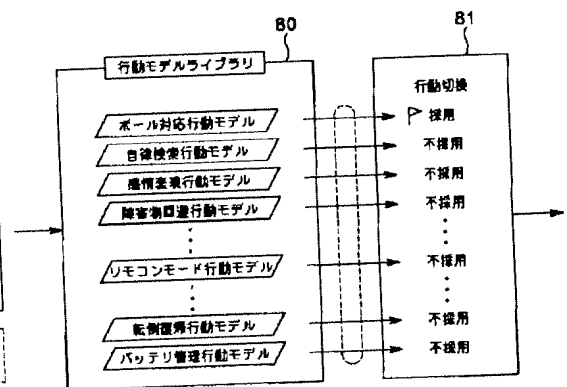
【図8】



【図9】



【図10】



【図12】

	入力イベント名	データ名	データの値	他のノードへの遷移確率				n
				A	B	C	D	
node 100								node 800
遷移先ノード				node 120	node 120	node 1000		
出力行動				ACTION 1	ACTION 2	MOVE BACK		ACTION 4
1	BALL	SIZE	0.1000	30%				
2	PAT				40%			
3	HIT				20%			
4	MOTION					50%		
5	OBSTACLE	DISTANCE	0.100			100%		
6		JOY	50.100					
7		SURPRISE	50.100					
8		SADNESS	50.100					

90

フロントページの続き

(51)Int.Cl.⁷
G 1 0 L 21/04

識別記号

F I
G 1 0 L 3/02
5/04

テ-マコード(参考)

A
F

(72)発明者 赤羽 誠
東京都品川区北品川6丁目7番35号 ソニ
ー株式会社内

Fターム(参考) 2C150 CA01 DA02 DF02 DF33 DG13
EB01 ED42 ED56 ED67 EE02
EE07 EF16 EH07 FA04
5D045 AA07 AB30